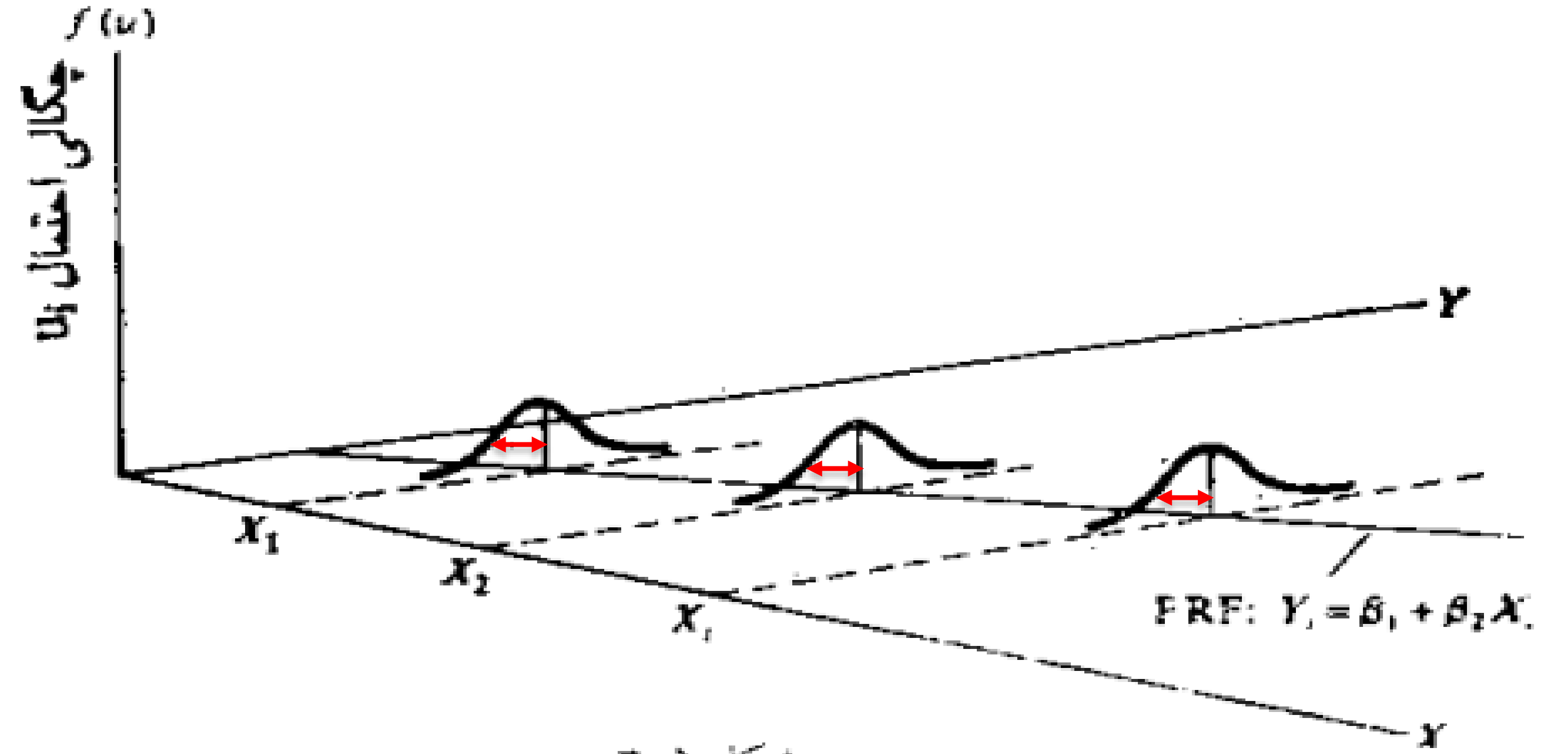
تهیه کننده: حسین خاندانی
Econometrics.blog.ir

بسم الله الرحمن الرحیم

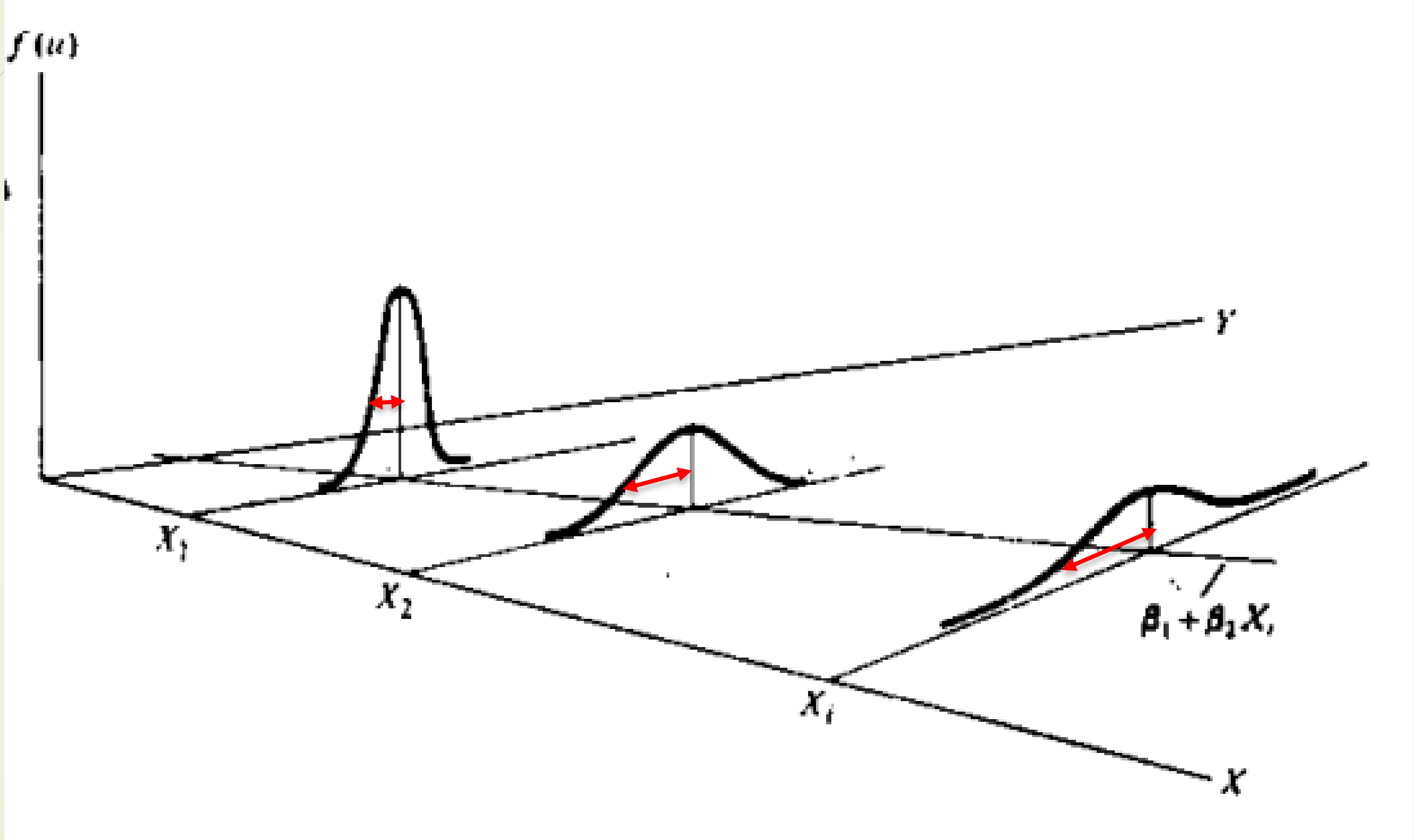
آموزش روش رفع ناهمسانی واریانس
EGLS یا FGLS

همسانی واریانس



شکل ۵-۳

ناهمسانی واریانس



دسته بندی واریانس رگرسیون

1. واریانس بلندمدت (غیرشرطی)

به ازای متغیرهای توضیحی

2. واریانس کوتاه مدت (شرطی)

در طی زمان با وقفه های خود

انواع واریانس شرطی (کوتاه مدت)

الف) واریانس ناهسانی به شرط متغیرهای توضیحی:

1) آزمون ناهمسانی واریانس بروش پاگان

2) آزمون ناهمسانی واریانس وایت (فقط روی پسماندهای حاصل از LS)

3) و ...

ب) واریانس ناهمسانی طی زمان و به شرط وقفه های خود متغیر:

۱) آزمون ناهسانی واریانس شرطی آرچ **(only time series)**

اصولا زمانی ما می گوییم واریانس شرطی منظور همان نسبت به زمان (اثر آرچ) است و نه به شرط متغیرهای مستقل.

نحوه انجام آزمون

برای انجام هر یک از آزمون های فوق ابتدا رگرسیون اصلی را برآورد و سپس پسماندهای مدل را ذخیره نموده و آنها را به توان رساند و به عنوان متغیر وابسته و نماینده واریانس در نظر گرفت.

$$y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i$$

$$\downarrow$$

$$\widehat{U}^2 = e^2$$

در آمار واریانس را برای متغیر مشاهده شده به صورت زیر بدست می آوردیم:

$$\mathrm{Var(x)}=\frac{\sum(X_i-\bar{X})^2}{obs}$$

چون **U** غیر قابل مشاهده است ما از امید ریاضی استفاده مینماییم و انتظار داریم واریانس جز خطا برابر با مجذور آن باشد.

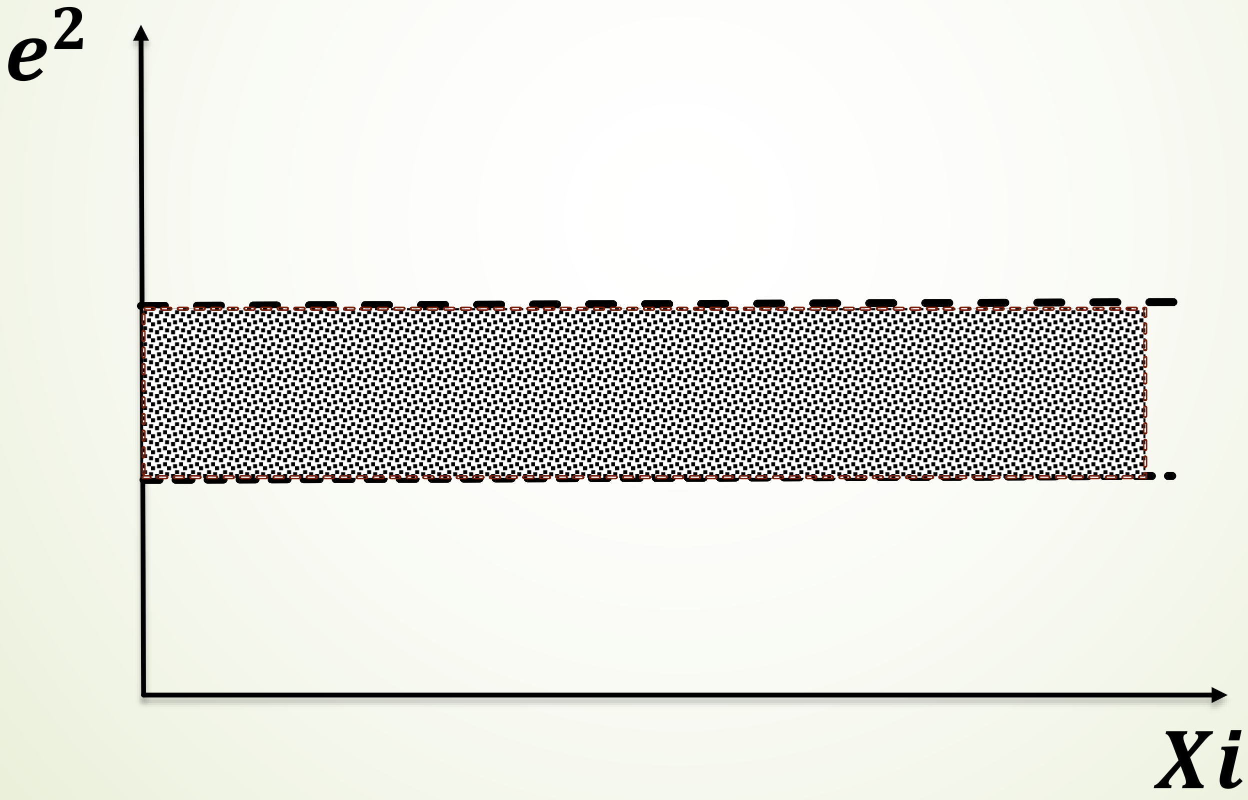
$$VAR(u_i)=E(u-E(u))^2=E(u^2)$$

اما به محض اینکه یک پدیده اتفاق افتاد دیگر ما از امید استفاده نمی نماییم. امید ریاضی تقریبا مفهومی برابر میانگین (انتظار)دارد در نتیجه خواهیم داشت.

$$Mean(X)=\frac{\sum X_i}{obs}=\frac{\sum e_i^2}{n-k}=\frac{RSS}{n-k}=\sigma^2=Var(u_i).$$

روش تشخیص ناهسانی واریانس به صورت ترسیمی

واریانس **هیچ ارتباطی** با متغیرهای توضیحی ندارد و همواره با تغییر متغیرهای توضیحی نوسانات **ثابتی** دارد.



آزمون بروش پاگان

آزمون بروش-پاگان به منظور آزمودن واریانس ناهمسانی در مدل‌های رگرسیون خطی استفاده می شود و وابستگی واریانس جملات پسماند بدست آمده از رگرسیون خطی را به مقادیر متغیرهای توضیح دهنده مدل، بررسی می کند.

$$u_i^2 = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \ldots + \gamma_k X_{ki} + \eta_i$$

آزمون وایت

آزمون وایت حالت کلی تر بروش پاگان است و نسبت به تشخیص واریانس ناهمسانی حساس تر است.

$$e_i^2 = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + \gamma_3 X_{1i}^2 + \gamma_4 X_{2i}^2 + \gamma_5 X_{1i} X_{2i} + v_i$$

اگر تعداد متغیرهای توضیحی را در تخمین بیش از اندازه باشند. دچار مشکل هم خطی می‌شویم بنابراین در تشخیص ناهمسانی واریانس دچار اشتباه شویم.

همچنین ممکن است درجه آزادی زیادی را از دست دهیم و دیگر قادر به تخمین مدل نباشیم (میتوان از وایت تعدیل شده استفاده نمود).

آزمون ناهمسانی واریانس شرطی آرچ

این آزمون تنها برای مدلهای کاربرد دارد که ما بعد زمان را داریم و در داده های مقطعی کاربردی ندارد.

$$e_t^2 = \alpha + \underbrace{\beta_1 e_{t-1}^2 + \beta_1 e_{t-2}^2 + \ldots +}_{\text{متغیر توضیحی}} \varepsilon_t$$

مشکل ناهمسانی واریانس و روش رفع آن

در حالت ناهمسانی واریانس دیگر واریانس یک مقدار ثابت نیست و همواره در هر مشاهده تغییر میکند (σ_i) . (فرمول واریانس اشتباه تخمین میخورد)

در حالت ناهمسان	در حالت همسان

$$\sum = E(uu') = \sigma^2 I \longrightarrow |\sum| = E(uu') = \sigma_1^2\sigma_2^2\ldots\sigma_n^2$$

$$\operatorname{var}(\hat{b}) = \frac{\sigma^2}{\sum x_i^2} \longrightarrow \operatorname{var}(\hat{b}) = \frac{\sum x_i^2 \sigma_i 2}{\left(\sum x_i^2\right)^2} \xrightarrow{\text{Robust}} \text{white}$$

سوال: ناهمسانی واریانس باعث چه مشکلاتی میشود؟

1. ناهمسانی واریانس ویژگی بدون تورش بودن تخمین های OLS را تغییر نمی دهد.
2. ناهمسانی ویژگی سازگاری تخمین زننده های OLS را نیز تغییر نمی دهد.
3. در حالت ناهمسانی واریانس تخمین زننده های OLS دیگر حداقل واریانس و یا کارایی را ندارند.

سوال: ناهمسانی واریانس باعث چه مشکلاتی میشود؟

1. ناهمسانی واریانس ویژگی بدون تورش بودن تخمین های OLS را تغییر نمی دهد. (چرا؟؟؟؟)
2. ناهمسانی ویژگی سازگاری تخمین زننده های OLS را نیز تغییر نمی دهد.
3. در حالت ناهمسانی واریانس تخمین زننده های OLS دیگر حداقل واریانس و یا کارایی را ندارند.

از آنجا که تخمین زننده ها سازگار و بدون تورش اند، بنابراین مشکل ناهمسانی واریانس چیست؟ و یا عدم کارایی باعث چه مشکلی میشود؟

سوال: ناهمسانی واریانس باعث چه مشکلاتی میشود؟

1. ناهمسانی واریانس ویژگی بدون تورش بودن تخمین های OLS را تغییر نمی دهد. (چرا؟؟؟؟)
2. ناهمسانی ویژگی سازگاری تخمین زننده های OLS را نیز تغییر نمی دهد. $E(\hat{B}) = B + (x'x)^{-1}E(x'u)$
3. در حالت ناهمسانی واریانس تخمین زننده های OLS دیگر حداقل واریانس و یا کارایی را ندارند.

از آنجا که تخمین زننده ها سازگار و بدون تورش اند، بنابراین مشکل ناهمسانی واریانس چیست؟ و یا عدم کارایی باعث چه مشکلی میشود؟

*زمانی که واریانس ها حداقل یا کارآ نباشند بنابراین فواصل اطمینان قابل اعتماد نیست و در نتیجه ما نمیتوانیم استنباط آماری انجام دهیم و آزمون فرضیه مخدوش میشود. **(F, T,…)**

در حالت ناهمسانی واریانس **GLS** چیزی نیست جز همان روش **WLS**

به طور خلاصه میتوان گفت روش gls با دادن وزن معکوس واریانس به متغیرها باعث میشود مشاهداتی که پراکندی بیشتری دارند وزن کمتر و مشاهداتی که پراکندگی کمتری دارند وزن بیشتری بگیرند و این مشاهدات در رگرسیون موثرتر واقع شوند.

$$y_i = B_1 + B_2 X_i + u_i$$

در حالتی که واریانسهای ناهمسان معلوم باشد به صورت زیر عمل میکنیم

$$\frac{y_i}{\sigma_i} = B_1\left(\frac{1}{\sigma_i}\right) + B_2\frac{X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$$

مدل تعمیم یافته:

$$y_i^* = B_1^* + B_2\ X_i^* + v_i^*$$

$$\operatorname{var}(v_i^*) = E(v_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 = \frac{1}{\sigma_i^2}E(u_i^2) = \frac{1}{\sigma_i^2}(\sigma_i^2) = 1$$

FGLS

بنابراین با فرض معلوم بودن $\sum .$، تخمین زننده های GLS قابل کاربرد است. اما اگر $\sum .$ معلوم نباشد. آنگاه بایستی ابتدا $\sum .$ را برآورد نموده و سپس مدل را تعدیل کرد.

اگر واریانس ناهمسان باشد باید پی برد که واریانس ناهمسان با چه شکل و تابعی نسبت به متغیرهای مستقل در حال تغییر می باشد:

$$\sum = \sigma_i^2 = \sigma^2 f(x) = \sigma^2 x^h$$

رفع مشکل ناهمسانی واریانس هنگامی که σ_i مجهول است

فرض 1: اگر فرض کنیم که واریانس U متناسب با مربع متغیر توضیحی X است

$$E(u_i^2) = \sigma_i^2 = \sigma^2 X_i^2$$

$$\frac{y_i}{X_i} = \frac{B_1}{X_i} + B_2 \frac{X_j}{X_i} + \frac{u_i}{X_i} \quad = B_1^* + B_2 X_i^* + \boldsymbol{v_i^*}$$

$$E(u_i^2) = \sigma_i^2 = \sigma^2 X_i^2$$

$$\operatorname{var}(v_i^*) = E(v_i^*)^2 = E\left(\frac{u_i}{X_i}\right)^2 = \frac{1}{X_i^2} E(u_i^2) = \frac{1}{X_i^2}(\sigma^2 X_i^2) = \sigma^2$$

رفع مشکل ناهمسانی واریانس هنگامی که σ_i مجهول است

فرض 2: اگر معتقدم باشیم که واریانس u، متناسب با خود متغیر توضیحی X تغیر میکند

$$E(u_i^2) = \sigma^2 X_i$$

$$\frac{y_i}{\sqrt{X_i}} = \frac{B_1}{\sqrt{X_i}} + B2\frac{X_i}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}}$$

$$\mathrm{Y}^* = B_1^* + B_2 X_i^* + \boldsymbol{v}_{\boldsymbol{i}}^*$$

$$\mathbf{var}(\boldsymbol{v}_{\boldsymbol{i}}^*) = \boldsymbol{E}(\boldsymbol{v}_{\boldsymbol{i}}^*)^2 = \frac{\mathbf{1}}{\boldsymbol{X}_{\boldsymbol{i}}}\boldsymbol{E}(\boldsymbol{u}_{\boldsymbol{i}}^2) = \frac{\mathbf{1}}{\boldsymbol{X}_{\boldsymbol{i}}}(\boldsymbol{\sigma}^2 \boldsymbol{X}_{\boldsymbol{i}}) = \boldsymbol{\sigma}^2$$

رفع مشکل ناهمسانی واریانس هنگامی که σ_i مجهول است

فرض 3: اگر معتقدم باشیم که واریانس u، متناسب با مربع امید ریاضی Y تغیر میکند

$$E(u_i^2) = \sigma^2 [E(Y_i)]^2$$

$$\boldsymbol{E(Y_i) = B1 + B2X_i}$$

$$\frac{y_i}{E(y)} = \frac{B_1}{E(y)} + B2\frac{X_i}{E(y)} + \frac{u_i}{E(y)}$$

$$\frac{Y_i}{\hat{Y}_i} = B1\left(\frac{1}{\hat{Y}_i}\right) + B2\left(\frac{X_i}{\hat{Y}_i}\right) + v_i$$

۴. تبدیل لگاریتمی:

تبدیل لگاریتمی غالبا مشکل ناهمسانی واریانس را کاهش میدهد، به دلیل اینکه مقیاس های اندازه گیری متغیر را تغییر میدهد؛ در نتیجه هر ۱۰ اختلاف بین دو متغیر به دو اختلاف کاهش می یابد

FGLS

حداقل مربعات تعمیم یافته شدنی FGLS

اصولا اطلاعات در مورد اینکه واریانس U، با چه تناسبی نسب متغیرهای توضیحی تغییر میکند بسیار ناقص و اندك است؛ و روش های فوق زمانی کاربرد دارد که این اشکال و یا ماتریس (Ω یا $\sum$) **معلوم** باشد.

اما این شرط در دنیای واقع به ندرت صادق است و بدست آوردن یک برآورد حداقل مربعات تعمیم یافته شدنی (FGLS)که در آن پارامترهای مجهول جای خود را به برآوردهای سازگار میدهند اهمیت پیدا میکند.

FGLS

حالت کلی:

$$Var(\varepsilon_i) = \sigma^2 X_i^h$$

مثال:

$$Var(\varepsilon_i) = \sigma^2 X_i^2 \ (h = 2) \longrightarrow \text{وزن} \ \boldsymbol{w = 1/x_i}$$

$$Var(\varepsilon_i) = \sigma^2 X_i (h = 1) \longrightarrow \boldsymbol{w = 1/\sqrt{x_i} = 1/x_i^{0.5}}$$

FGLS

فرمول واریانس

$$Var(e_i) = E(e^2) = \sigma^2 X_i^h$$

FGLS

مجهول؟؟

$$Var(e_i) = E(e^2) = \sigma^2 X_i^h$$

حال از دو طرف لگاریتم بگیرید

$$\ln(e_i^2) = \underbrace{\ln(\sigma^2)}_{\text{عرض از مبدا}=a} + h\ln(X_i) + \zeta_i$$

بنابراین شیبی که بدست می آید همان وزن و شکل واریانس ناهمسانی است

FGLS

* گام های زیر را انجام دهید:

1. Estimate the regression with OLS.
2. Regress

$$\ln(e_i^2) = \ln(\sigma^2) + h\ln(X_i) + \zeta_i$$

3. Divide every variable by:

$$d_i = \sqrt{X_i^{\hat{h}}} = X_i^{\frac{\hat{h}}{2}}$$

4. Apply OLS to the transformed data.

نحوه انجام در نرم افزار

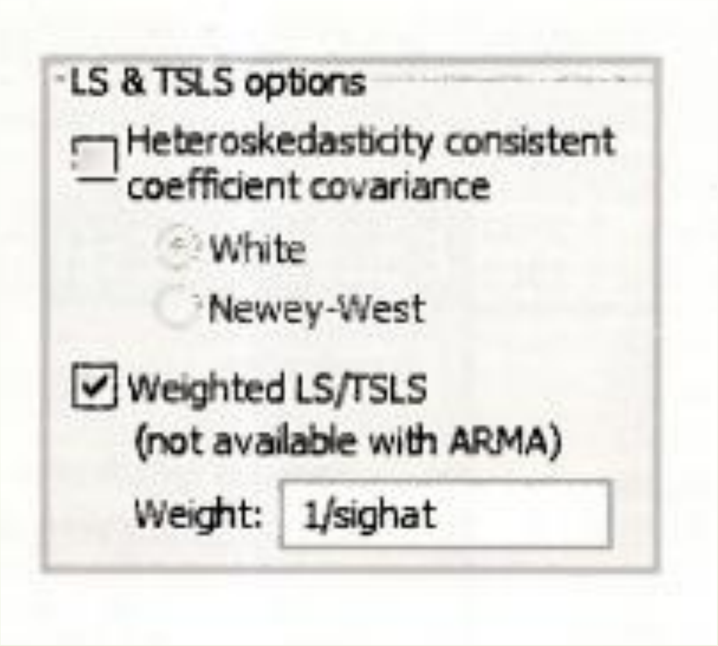
پسماند را بتوان ۲ رسانده و جایگزین متغیر وابسته نموده و مدل را بصورت لگاریتمی تخمین می زنیم. ضریب شیبی که بدست آمده است با تقسیم بر ۲ همان شکل واریانس ناهمسانی و یا توان وزن را نشان میدهد. البته متدهای دیگر شبیه به این روش وجود دارد که میتوان از آنها نیز استفاده نمود معادله واریانس نیمه لگاریتمی.

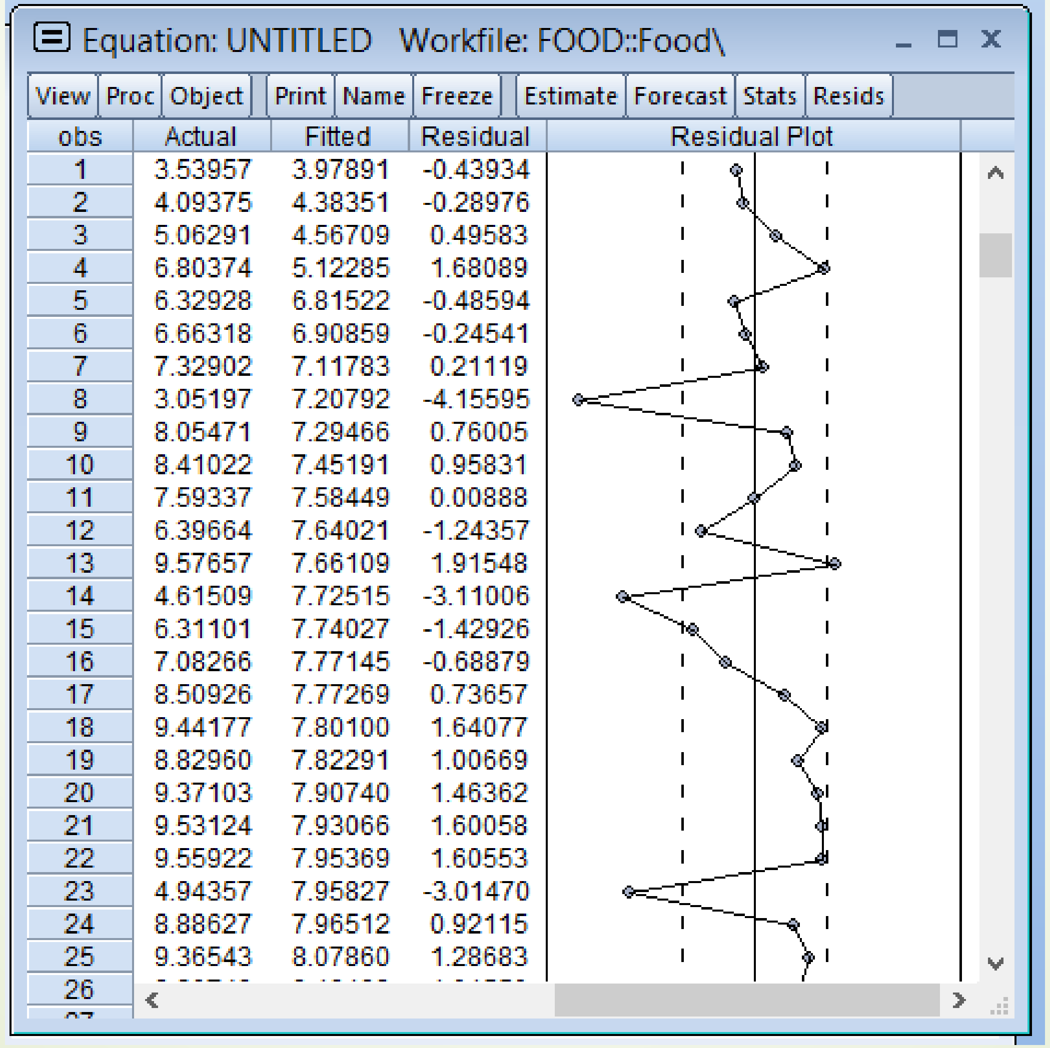
اما زمانی که بیش از یک متغیر مستقل وجود دارد بهتر است از روش زیر استفاده شود.

یک روش کلی در زمانی که بیش از یک متغیر مستقل وجود دارد

To proceed to generalized least squares estimation we need the exponential of the predictions from this equation, $\hat{\sigma}_i^2 = \exp\left(\hat{\alpha}_1 + \hat{\alpha}_2 \ln(INCOME_i)\right)$ or their square roots $\hat{\sigma}_i$. It is instructive to consider two ways of computing them. The first is using the commands

```
series  sig2hat = exp(c(1) + c(2)*log(income))
series  sighat = @sqrt(sig2hat)
```





Equation: UNTITLED   Workfile: FOOD::Food\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

obs	Actual	Fitted	Residual	Residual Plot
1	3.53957	3.97891	-0.43934	
2	4.09375	4.38351	-0.28976	
3	5.06291	4.56709	0.49583	
4	6.80374	5.12285	1.68089	
5	6.32928	6.81522	-0.48594	
6	6.66318	6.90859	-0.24541	
7	7.32902	7.11783	0.21119	
8	3.05197	7.20792	-4.15595	
9	8.05471	7.29466	0.76005	
10	8.41022	7.45191	0.95831	
11	7.59337	7.58449	0.00888	
12	6.39664	7.64021	-1.24357	
13	9.57657	7.66109	1.91548	
14	4.61509	7.72515	-3.11006	
15	6.31101	7.74027	-1.42926	
16	7.08266	7.77145	-0.68879	
17	8.50926	7.77269	0.73657	
18	9.44177	7.80100	1.64077	
19	8.82960	7.82291	1.00669	
20	9.37103	7.90740	1.46362	
21	9.53124	7.93066	1.60058	
22	9.55922	7.95369	1.60553	
23	4.94357	7.95827	-3.01470	
24	8.88627	7.96512	0.92115	
25	9.36543	8.07860	1.28683	
26	[illegible]	[illegible]	[illegible]	

$$Y = E(Y|X) + U_i = \hat{Y} + e_i$$

$$e_i^2 = E(e^2|x_i) + v_i$$

$$e_i^2 = \sigma_i^2 + v_i$$

Actual **Fitted** **Residual**

$$\ln(e_i^{\,2}) = \ln(\sigma^2) + h\ln(X_i) + \zeta_i$$

قسمت ناهمسان واریانس

$ln(\sigma i^2)$

لگاریتم واریانس خالص شده از متغیرهای مستقل

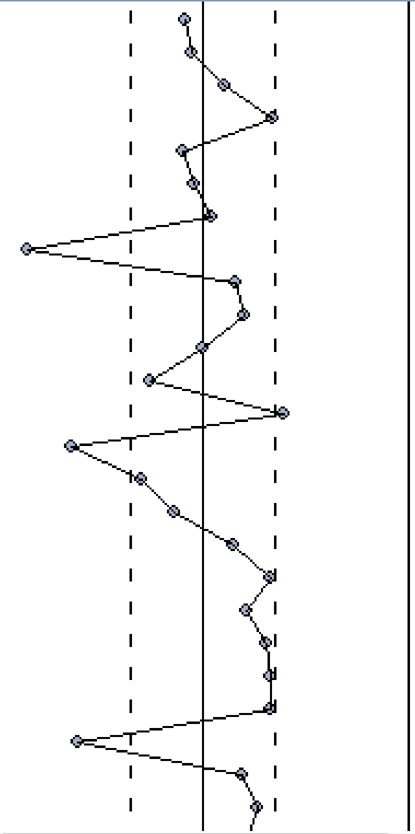
$v_i = ln(\sigma^2)$

$ln(ei^2)$

Equation: UNTITLED Workfile: FOOD::Food\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

obs	Actual	Fitted	Residual	Residual Plot
1	3.53957	3.97891	-0.43934	
2	4.09375	4.38351	-0.28976	
3	5.06291	4.56709	0.49583	
4	6.80374	5.12285	1.68089	
5	6.32928	6.81522	-0.48594	
6	6.66318	6.90859	-0.24541	
7	7.32902	7.11783	0.21119	
8	3.05197	7.20792	-4.15595	
9	8.05471	7.29466	0.76005	
10	8.41022	7.45191	0.95831	
11	7.59337	7.58449	0.00888	
12	6.39664	7.64021	-1.24357	
13	9.57657	7.66109	1.91548	
14	4.61509	7.72515	-3.11006	
15	6.31101	7.74027	-1.42926	
16	7.08266	7.77145	-0.68879	
17	8.50926	7.77269	0.73657	
18	9.44177	7.80100	1.64077	
19	8.82960	7.82291	1.00669	
20	9.37103	7.90740	1.46362	
21	9.53124	7.93066	1.60058	
22	9.55922	7.95369	1.60553	
23	4.94357	7.95827	-3.01470	
24	8.88627	7.96512	0.92115	
25	9.36543	8.07860	1.28683	
26				



$$\boldsymbol{Y = E(Y|X) + U_i = \hat{Y} + e_i}$$

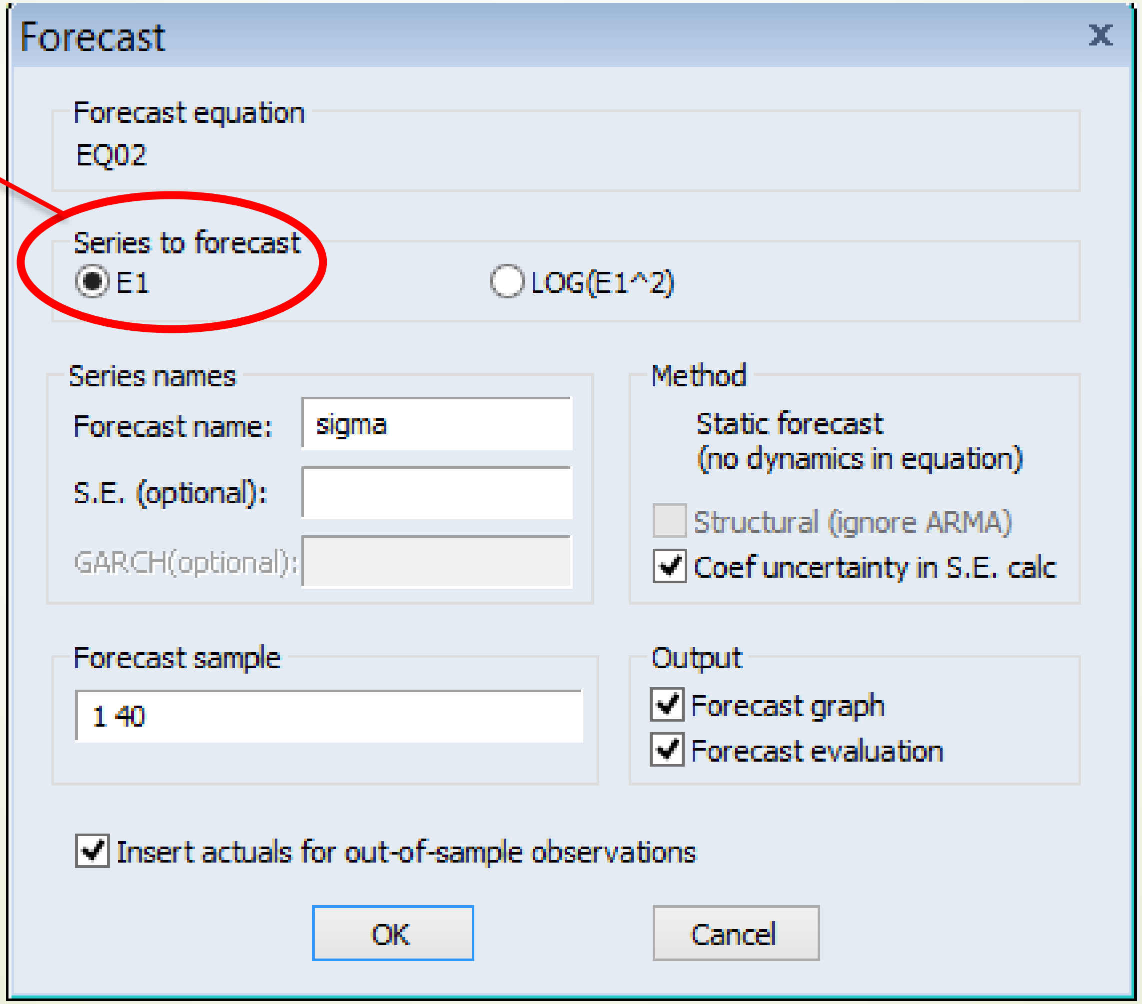
$$\boldsymbol{e_i^2 = E(e^2|x_i) + v_i}$$

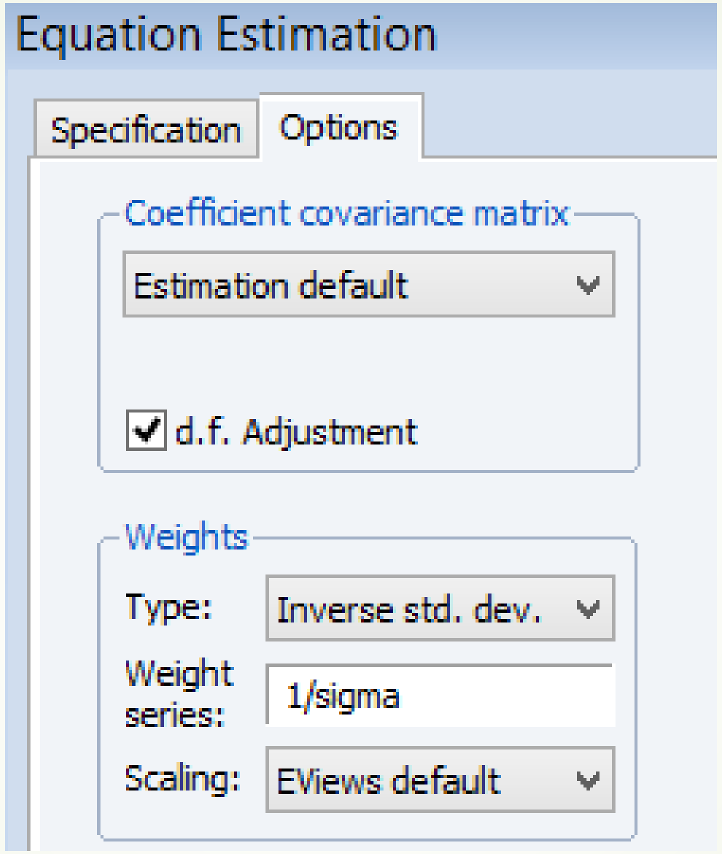
$$\boldsymbol{e_i^2 = \sigma_i^2 + v_i}$$

Actual **Fitted** **Residual**

روش میان بر از طریق دستور Forecast

Sigma




Equation Estimation
Specification
Options
Coefficient covariance matrix
Estimation default
d.f. Adjustment
Weights
Type:
Inverse std. dev.
Weight series:
1/sigma
Scaling:
EViews default

Equation: EQ01 Workfile: FOOD::Food\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: White

F-statistic	0.509466	Prob. F(3,36)	0.6783
Obs*R-squared	1.629057	Prob. Chi-Square(3)	0.6528
Scaled explained SS	0.830585	Prob. Chi-Square(3)	0.8421

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.108747	Prob. F(1,38)	0.2990
Obs*R-squared	1.134014	Prob. Chi-Square(1)	0.2869
Scaled explained SS	0.578184	Prob. Chi-Square(1)	0.4470

تهیه کننده: حسین خاندانی
Econometrics.blog.ir

زکات علم نشر آن است